

CHAPTER I

INTRODUCTION

Concern about the quality and effectiveness of American elementary and secondary education has been unusually intense for several years, perhaps greater than at any time since the Sputnik-inspired reform era of three decades ago. This concern has had many expressions: extensive coverage in the press, numerous influential reports on the status of education, and widespread political attention and efforts at all levels of government to improve the educational system.

Measures of educational achievement, particularly scores on various types of standardized tests, have played a key role in this ferment. One of the wellsprings of the debate was a growing public awareness that by many measures, the educational achievement of American students dropped considerably during the 1960s and 1970s, and that it compares unfavorably with the performance of students in some other countries. This information from educational tests, and the abundant hypotheses about the causes of these deficiencies in performance, have played a central role in forming the current spate of educational initiatives at all levels of government. Many of these initiatives are responses to problems revealed by such tests, and test scores have been cited as being a part of their rationale.

The influence of tests on both educational practice and public discussion has also increased. Many of the recent educational initiatives entail using these tests more and giving them greater importance. Examples include increased reliance on tests as prerequisites for high school graduation and the use of tests to screen potential teachers. In addition, Americans appear to have come increasingly to judge the quality of their schools by the results of achievement tests--a trend that is apparent from the local level to the national. Indeed, standardized tests have become a sort of national report card. Local newspapers routinely publish comparisons of schools in terms of the average test scores of their students. At the national level, the Department of Education has begun publishing periodic comparisons of the educational systems of the 50 states, highlighting the average scores on college admissions tests of the students in each state who take those tests.

Over the past year or so, more positive trends in educational achievement have gained increasing attention. It is now widely known that the decline of test scores during the 1960s and 1970s has ended and has been followed by a substantial rise. Although the more favorable recent trends in test scores have not yet affected the current wave of educational policy initiatives in a way comparable to that of the preceding decline, they too have been incorporated into the national report card and have been cited by many observers as an indication that the educational system is improving.

The current importance attached to test data makes it critical to appraise recent trends in test scores accurately and to evaluate explanations of those trends carefully. *Trends in Educational Achievement*, a Congressional Budget Office study released in April 1986, assessed much of the available information about trends in test scores and described some of the important characteristics and limitations of common tests. (Several conclusions of the earlier report that are crucial to an understanding of this paper are summarized here in Chapter II.) This report supplements the earlier one by analyzing possible causes of trends in test scores. Some of the most common or influential explanations are evaluated by assessing their consistency with the broad array of test data analyzed in the earlier report and with other, independent evidence. In addition, this report explores the implications for policy of both the trends and their causes.

THE CONTEXT OF THE CURRENT CONTROVERSY

While elementary and secondary education remains primarily a state and local responsibility, it is a truly national concern. Debate about education policy frequently stresses questions of national interest, such as the impact of education on the productivity of the nation's work force and consequently on the international competitiveness of the American economy and the nation's security. The current debate has been shaped by the reports of numerous national commissions, the National Governors' Association, the Council of Chief State School Officers, and the Department of Education, as well as other regional and national groups. Moreover, many of the recent changes in educational policy and practice have been national in scope, as many states followed common paths in making independent decisions.

The themes of the current controversy--and participation in the debate by members of the Congress and the Administration--have long-standing historical precedents. This continuity is perhaps clearest in the concern about the possible consequences of education for the productivity of the work force and the competitiveness of the American economy, which

has been a recurring theme in legislation and in debates about educational policy at least since the turn of the century. For example, one of the aims of the Smith-Hughes Act of 1917, which established federal support for vocational education, was to improve the skills and productivity of the work force in response to international competition.^{1/} That act, which is commonly acknowledged as the first federal program of categorical aid to elementary and secondary education, is still funded today.

More recently, the report of the National Commission on Excellence in Education, *A Nation at Risk*, stated that "Our once unchallenged preeminence in commerce, science, and technological innovation is being overtaken by competitors throughout the world. This report is concerned with only one of the many causes and dimensions of the problem, but it is the one that undergirds American prosperity, security, and civility."^{2/} A particularly influential report, *A Nation Prepared: Teachers for the 21st Century*, issued by the Carnegie Forum on Education and the Economy, asserted that "America's ability to compete in world markets is eroding...As in past economic and social crises, Americans turn to education. They rightly demand an improved supply of young people with the knowledge...and skills to make the nation once again fully competitive."^{3/}

Concern has also been voiced about the perceived failure of the educational system to challenge the nation's most able students. This too has been a recurrent theme and can be traced back at least as far as the 1893 report of the "Committee of Ten," considered by some historians to be the first major national report on the high school. This concern has been the focus of several recent congressional initiatives.

THE FEDERAL ROLE IN ELEMENTARY AND SECONDARY EDUCATION

The federal government has always played a more limited role in elementary and secondary education than have states and localities. Together, states

-
1. Carl F. Kaestle and Marshall S. Smith, "The Federal Role in Elementary and Secondary Education, 1940-1980," *Harvard Educational Review*, vol. 54, no. 4 (November 1982), pp. 384-408.
 2. National Commission on Excellence in Education, *A Nation at Risk* (Washington, D.C.: Government Printing Office, 1983), p. 5.
 3. Task Force on Teaching as a Profession, *A Nation Prepared: Teachers for the 21st Century* (Washington, D.C.: Carnegie Forum on Education and the Economy, May 1986), p. 2.

and localities provide most of the funds for public education--over 90 percent, by the most common accounting--and they retain control over most aspects of educational policy and practice.^{4/} Decisions about teacher certification, curricula and course requirements, and achievement testing, for example, all rest with state and local governments.

Nonetheless, the roles of the Congress and the Administration have at times been more significant than the relatively small federal share of funding might suggest. In certain areas, such as the education of handicapped or educationally disadvantaged students, the federal role is central. The federal government also influences elementary and secondary education by means other than the funding of educational services; it assumes major responsibility for collecting and disseminating educational information and statistics.

Changes in the Scope of Federal Aid to Education

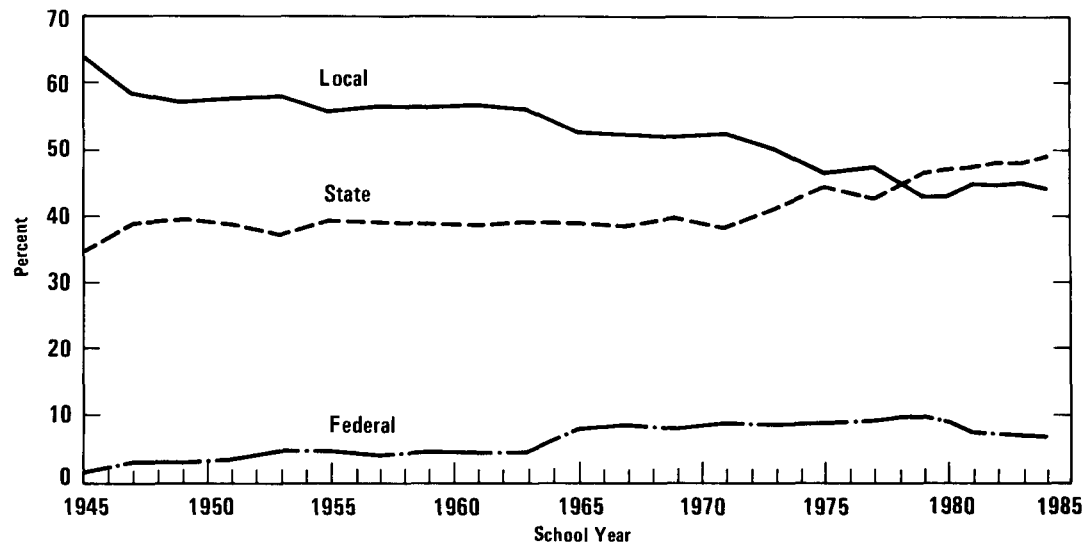
During the decades following World War II, federal aid to education grew markedly. Until the mid-1940s, federal contributions had accounted for less than 1.5 percent of total revenues for public elementary and secondary education. The federal share then rose markedly for about three decades, reaching a peak of almost 10 percent in the late 1970s (see Figure 1). Since then, the federal share has fallen considerably. In the 1984-1985 school year, federal contributions of nearly \$9 billion constituted about 6.5 percent of total revenues for education--the smallest share in two decades.

The postwar increase in the federal share of education revenues reflected major qualitative changes in the goals of federal involvement. Until the 1950s, federal aid for education was devoted to only a few purposes, such as vocational education, the education of Native American children, and fiscal assistance to localities affected by federal installations. Moreover, in 1950, more than half of all federal aid was provided for the school lunch program, not for specifically educational programs.

Since 1950, a variety of laws have broadened the scope of federal assistance for education. The National Defense Education Act of 1958 (NDEA), for example, authorized various activities intended to improve instruction in mathematics, sciences, and foreign languages. The Elementary and Secondary Education Act of 1965 (ESEA, Public Law 89-10), which produced the large increase in federal funding in the mid-1960s, authorized

4. Department of Education, Office of Educational Research and Improvement, *Digest of Education Statistics, 1985-86* (February 1986), Table 69.

Figure 1.
Shares of Elementary/Secondary Education Funding
by Level of Government



SOURCE: Office of Educational Research and Improvement, *Digest of Education Statistics, 1987* (Washington, D.C.: Department of Education, 1987).

a wide range of programs, including the program of compensatory education that--as Chapter 1 of the Education Consolidation and Improvement Act of 1981--remains the largest single source of federal funds for elementary and secondary education.

Although these programs represented substantive changes in the character of federal aid, many of the rationales behind them echoed earlier concerns. The statement of the purpose of the NDEA, for example, noted that:

The Congress hereby finds and declares that the security of the Nation requires the fullest development of the mental resources and technical skills of its young men and women. The present emergency demands that additional and more adequate educational opportunities be made available. . . . 5

5. Public Law 85-864, Section 101 (72 Stat. 1580). See also Kaestle and Smith, "The Federal Role in Elementary and Secondary Education," p. 393.

Similarly, although the main purpose of the ESEA was to improve the opportunities open to disadvantaged students, it too reflected the concerns of Smith-Hughes and the NDEA--the effect of inadequate education on the nation's well-being. 6/

Federal Support of Educational Statistics and Research

In addition to providing financial support for certain educational services, the federal government has long been involved in elementary and secondary education by generating, collecting, and disseminating statistics and research about education. The U.S. Department of Education was established in 1867 primarily to gather educational statistics, and that function has continued without interruption to the present. The Bureau of the Census also collects statistical information about students and school districts.

This role has grown substantially in recent years. The Education Amendments of 1972 (Public Law 92-318), for example, established the National Institute of Education, now a part of the Office of Educational Research and Improvement, which has been a major source of funding for research on education. Federal efforts to gather or disseminate educational information have also accompanied programs of direct financial support of educational services. Several current proposals would further expand the federal role in gathering educational information. The report of the Secretary of Education's panel on improving the assessment of student performance, for example, recommended greatly expanding the National Assessment of Educational Progress to permit state-by-state comparisons of student achievement. 7/

Although information-related activities absorb only a modest share of federal funding for elementary and secondary education, the federal funds provide a large part of the resources for carrying them out. 8/ In a number of cases, the data generated by the federal government are unique. For

-
6. *Elementary and Secondary Education Act of 1965*, Report No. 89-143, House Committee on Education and Labor, to accompany H. R. 2362, 89:1 (1965), pp. 1448-1449.
 7. Lamar Alexander, H. Thomas James, and others, *The Nation's Report Card: Improving the Assessment of Student Achievement* (Washington, D.C.: Office of Educational Research and Improvement, 1987).
 8. For example, in fiscal year 1986, funding for the Office of Educational Research and Improvement, which accounts for a large share of federal support for educational statistics and research, totaled about \$64 million--about three-tenths of one percent of the Education Department's appropriation of \$19.5 billion.

example, all of the nationally representative data on educational achievement test scores used in this and the preceding report were federally funded. Moreover, the impact of those data in many cases is far greater than their relatively small share of funding might suggest, because they can influence educational policy and practice at all levels of government.

RECENT POLICY INITIATIVES

The intensity of the current debate about educational achievement has been matched by the abundance of policy initiatives proposed--and, in many cases, already carried out--at all levels of government. Many states and localities have instituted sweeping policy changes affecting a wide range of educational practices. Common initiatives have included increased coursework requirements for high school graduation, expanded programs of student testing, changes in standards for teacher certification, and modifications of rules for teacher compensation.

The federal responses have also been diverse, but many have been consistent with past federal efforts. The Administration has emphasized its role of disseminating information in its attempts to alter education policy and practice--for example, by issuing comparisons of the states' educational policies and outcomes. Some of the legislation considered by the Congress has followed traditions established by the NDEA, the ESEA, and Smith-Hughes. The Education for Economic Security Act of 1984 (Public Law 98-377), for example, followed the path of the NDEA in attempting to strengthen instruction in mathematics and science. Provisions with similar goals are also included in the trade bills passed by both Houses during the first session of the 100th Congress and currently awaiting conference--H.R. 3, the Trade and International Economic Policy Reform Act of 1987, and S. 1420, the Omnibus Trade and Competitiveness Act of 1987. Following in the tradition of the ESEA were the Job Training Partnership Act Amendments of 1986 (Public Law 99-496), which required that remedial education be included in certain federally funded training programs, and S. 1420, which would provide funds for a secondary school basic skills program and a dropout prevention program. In the tradition established by Smith-Hughes, H.R. 3 would also provide additional support for vocational education.

Trends in educational achievement and their presumed causes have served as rationales for many of these initiatives. Some initiatives--for example, efforts to strengthen mathematics education--focus on areas in which students' performance has shown particularly serious weaknesses or

especially severe deterioration. Other initiatives, such as increases in graduation requirements, are intended to alter aspects of policy and practice that have been suggested as causes of the decline of the 1960s and 1970s, or to augment policies that might have contributed to the subsequent rise in scores.

Recent achievement trends represent only one basis for educational policy changes. Changing a particular practice might prove beneficial, for example, even if that practice--contrary to common views--did not contribute appreciably to the decline of test scores. For instance, the much discussed decline in the SAT scores between 1972 and 1979 of individuals expecting to become teachers occurred too late to have contributed appreciably to the decline in students' test scores, but that fact says nothing about the influence of teachers' academic skills on students' achievement more generally. Nonetheless, as long as the trends and their presumed causes are put forward as a justification of policy changes, it is important to evaluate the consistency between policies and these trends. Assuming greater consistency than actually exists can misdirect policy in numerous ways. It can lead to unwarranted presumptions about the effectiveness of policy initiatives, and it can obscure the importance of other factors that are less commonly viewed as being linked to the trends of the recent past.

CHAPTER II

EDUCATIONAL ACHIEVEMENT:

FACTS AND UNCERTAINTIES

Because many current educational initiatives are responses to recent trends in educational achievement or to their possible causes, it is crucial to understand what the available data indicate about the achievement of elementary and secondary school students. This chapter summarizes some of the most important patterns that emerge when a wide array of data about educational achievement is examined. It is largely adapted from *Trends in Educational Achievement*, which provides more detailed information and more fully explains the limitations of existing test data.

TEST SCORES AS A MEASURE OF EDUCATIONAL ACHIEVEMENT

Current data on educational achievement are more complex, varied, and ambiguous than many observers realize. That complexity alone signals a need for caution in reaching conclusions about the condition of education, in considering possible explanations of recent trends, and in drawing inferences about appropriate policy responses.

The current debate about educational achievement was sparked by and focuses primarily on the results of standardized tests, such as college admissions tests, minimum-competency tests, and "norm-referenced" tests (tests that rate students by comparing their performance to that of other students, rather than to an absolute criterion of achievement). The debate, in turn, has prompted the burgeoning use of tests and a reliance on their results as indicators of the condition of education. Given this pivotal role of standardized tests, the strengths and limitations of test scores as an indicator of achievement are critically important.

The advantages of certain tests are considerable and apparent. The scoring of standardized tests can be free of much of the subjectivity that plagues alternative measures, such as teachers' grades. If designed and scored appropriately, tests can provide information about changes in achievement over time. Tests can also be tailored to address a wide variety

of specific questions, such as the extent of progress among certain groups of students or in subject areas of particular importance.

The limitations of test scores, while less apparent, are also considerable and must be recognized. Perhaps most important, test scores are not synonymous with educational achievement; rather, a given test is usually only an incomplete proxy for the comprehensive measure of achievement that one would ideally want. Most tests can tap only a subset of the many, highly disparate skills subsumed by a subject area such as mathematics or American history. When the skills being tested are specific and narrowly defined--for example, facility with algorithms for subtraction--a test can be a reasonably close proxy. The concerns of educational policy are rarely that narrow, however. Policy debate is more likely to focus on mathematics, for example, than on subtraction. Assessing these broader areas of achievement forces important trade-offs in the design of tests. ^{1/}

In addition, some of the skills and attitudes that schools strive to foster are difficult to gauge using standardized tests, and the assessment of students' performance can be distorted by the scarcity of information about these characteristics in the available test data. For example, the ability to write cogently is hard to assess because evaluating writing samples is both laborious and subjective, particularly in comparison with multiple-choice tests. As a result, large-scale, direct assessments of writing ability (as opposed to multiple-choice tests of language usage and writing mechanics) have been relatively uncommon until recently and have had comparatively little influence on public perception of achievement trends. Other attributes that schooling attempts to develop may be even more difficult to assess, such as an interest in reading, mastery of certain types of reasoning, and the ability and propensity to apply skills developed in school to very different and perhaps unstructured problems encountered out of school.

Another limitation of test scores as an indicator of achievement is that even similar tests can yield markedly different results. Indeed, one of the most serious mistakes made by some analysts attempting to explain recent achievement trends--or to draw implications for policy--has been to assume that patterns evident in the scores of one test will appear in

-
1. Moreover, the range of subject matter need not be very broad to force important trade-offs. One recent study of fourth-grade mathematics, for example--a subject with relatively little curricular variation--found sizable differences in the content of commonly used tests. See Donald J. Freeman, Theresa M. Kuhs, Andrew C. Porter, Robert E. Floden, William H. Schmidt, and John R. Schwille, "Do Textbooks and Tests Define a National Curriculum in Elementary School Mathematics?" *The Elementary School Journal*, vol. 83, no. 5 (May 1983), pp. 501-513.

others as well. Some of the patterns that have been prominent in the recent debate about educational policy do not appear consistently when a wide array of tests are considered.

Given that tests are incomplete proxies for comprehensive measures of achievement, some discrepancies in their results should be expected, and some of the factors that contribute to the variation in results are known. Choices made in designing the tests, for example--decisions about content, emphasis, and test format--can cause the results of tests to vary. Results can also differ because of seemingly arcane technical details. For example, the answer to the key question of whether trends in achievement have been more favorable among low-achieving students than among their high-achieving peers varies depending on how the test scores are scaled and reported. Still, some important discrepancies in the results of major tests remain unexplained.

PATTERNS IN THE ACHIEVEMENT DATA

The available data from standardized tests paint a mixed picture of the achievement of elementary and secondary school students: some aspects of the data are encouraging, while others are profoundly disturbing. This duality is especially evident when one considers both the levels of achievement shown on various tests and the trends in achievement over time. For example, promising trends can appear even when average scores remain distressingly low.

The Decline in Test Scores

The sizable drop in test scores during the 1960s and 1970s is well known and need not be detailed here, but several aspects of that decline bear mention. Perhaps most important to an assessment of possible causes is the remarkable pervasiveness of the decline. The drop in test scores took place among many different types of students, in many subject areas, on diverse tests, in all parts of the nation, and in Catholic as well as public schools.^{2/} Indeed,

-
2. The achievement decline among private schools evident in nationally representative data largely reflects the drop in scores of students in Catholic schools; the data are insufficient to gauge separately the trends in non-Catholic private schools. See Donald Rock, Ruth B. Eckstrom, Margaret E. Goertz, Thomas L. Hilton, and Judith Pollack, *Factors Associated With Decline of Test Scores of High School Seniors, 1972 to 1980* (Washington, D.C.: Center for Statistics, Department of Education, 1985), Chapter 5 and Appendix D. This distinction between Catholic and other private schools was not noted in *Trends in Educational Achievement*, the report from which this chapter is adapted.

data on test scores from Canada, though limited, suggest that somewhat similar trends appeared there as well.^{3/} Available data do not pinpoint the onset of this decline precisely but suggest that it began in all affected age groups within a short period during the mid-1960s.

Though pervasive, the achievement decline showed substantial variations, and these variations--when they occur consistently in numerous tests--also shed light on possible causes. One of the most important of these differences is that the decline was greater among older students. The decline lasted longer in the higher grades; in addition, limited evidence suggests that scores dropped more rapidly on tests administered to older students, at least during the early years of declining scores. Thus, the tests that have received the greatest attention and that have shaped many observers' impressions of achievement trends--tests administered to high school students--generally showed the greatest drops in scores. In contrast, tests administered in the first three grades showed little or no decline, and those administered in the middle grades tended to show moderate declines.

Another, particularly distressing, variation in the data is that higher-order skills (that is, skills such as reasoning and problem-solving), which showed particularly severe weaknesses throughout the period considered, deteriorated more markedly in some instances than did the most basic skills (such as factual knowledge, literal decoding of written text, and mastery of computational algorithms). The National Assessment of Educational Progress, for example, found somewhat greater drops in performance in higher-order skills in both mathematics and reading.

The greater severity of the decline in scores in the upper grades might also be an indication of the sharper deterioration of higher-order skills, because the material included in tests administered in higher grades is progressively more complex. Indeed, the virtual absence of a decline in scores in the first three grades might partly reflect the emphasis on basic skills in tests administered in those grades. It is important to note, however, that the particularly severe problems with higher-order skills are also apparent even in the case of relatively simple material, including some taught in the elementary and junior-high grades. The National Assessment of Educational Progress, for example, found that large numbers of students

-
3. The primary source of data pertaining to Canadian students is from an adaptation of the Iowa Tests of Basic Skills administered to Canadian students through grade 8 in 1966, 1973, and 1980. See *Canadian Tests of Basic Skills: Manual for Administrators, Supervisors, and Counselors, Levels 5-18, Forms 5 and 6* (Scarborough, Ontario: Nelson Canada, 1984), p. 80; also, Thomas Schweitzer, Economic Council of Canada, personal communication, February 18, 1987.

are unable to apply basic arithmetic algorithms to the solution of simple word problems.

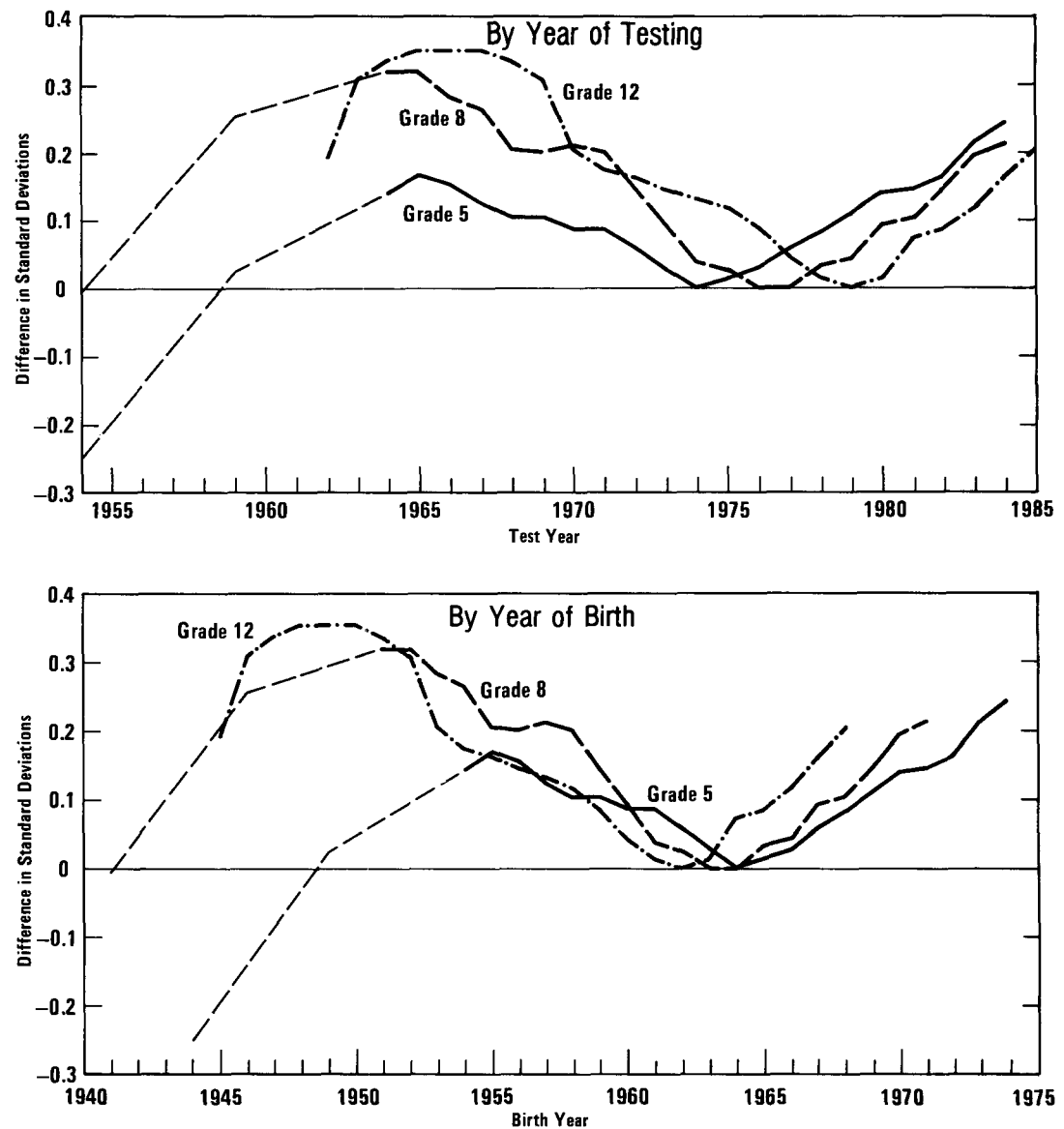
The Upturn in Test Scores

The current debate about education, while still shaped largely by the decline in test scores, has been altered recently by a growing awareness of favorable trends in achievement. It is now generally recognized that a widespread rise in test scores followed immediately on the heels of the decline and has been under way for some time. The characteristics of that upturn, however, are less well recognized. In particular, because of the greater attention afforded to tests administered at the high school level, such as the Scholastic Aptitude Test (SAT), many analysts have mistakenly believed that the rise in scores began within the past few years. In fact, the upturn was apparent in certain grades as early as the mid-1970s.

The upturn, like the preceding decline, was not uniform, and again variations in the trends hold keys to understanding their possible causes. Particularly important are differences among age groups. The decline in test scores ended--and the subsequent rise in scores began--first in the lower grades and later in the higher grades. The upturn first became apparent in test scores of students in the middle elementary grades in the mid-1970s. For example, in the Iowa state assessments--in some respects the best available data on trends in elementary and secondary achievement, although not representative of the nation as a whole--scores of fifth-grade students began climbing in 1975. The upturn then moved into the higher grades at a rate of roughly one grade per year, reaching the senior high school grades around the end of that decade. The end of the achievement decline and the onset of the following rise thus appear to constitute a "cohort effect"--a change that occurs in one or a few birth cohorts and therefore appears in different age groups as the affected cohorts grow older. This reversal in the trends occurred on most tests within a few years of the birth cohorts of 1962 and 1963 and moved up through the grades as those cohorts passed through school. (This pattern is clearest in the Iowa state data; see Figure 2.) Subsequent birth cohorts have typically scored progressively higher.

The upturn in scores in the lower grades has to date been larger than that in the upper grades. By some measures, the rise in achievement in the elementary grades has more than fully overcome the decline, so that scores are now at their highest point on record--a span of up to three decades. In contrast, scores on some tests administered in the higher grades remain considerably below their pre-decline high point. The greater improvement

Figure 2.
Iowa Average Test Scores, Grades 5, 8, and 12,
Differences from Post-1964 Low Point



SOURCES: Congressional Budget Office calculations based on "Iowa Basic Skills Testing Programs, Achievement Trends in Iowa: 1956-1985" (Iowa Testing Programs, unpublished and undated material); A.N. Hieronymus, E.F. Lindquist, and H.D. Hoover, *Iowa Tests of Basic Skills: Manual for School Administrators* (Chicago: Riverside, 1982); "Mean ITED Test Scores by Grade and Subtest for the State of Iowa, 1962 to Present" (Iowa Testing Programs, unpublished and undated tabulations); and Robert Forsyth, Iowa Testing Programs, personal communication, August 1984.

NOTE: Dashed lines (— —) indicate data available only at five-year intervals.

in the lower grades apparently has resulted largely from the longer duration of the rise in the lower grades--that is, the larger number of higher-performing birth cohorts who have so far reached the lower grades. The limited available data suggest that the annual rate of improvement has been roughly comparable in different grades (see Figure 2).

Variations in Trends Among Types of Students and Schools

Achievement trends have also varied among different groups of students. One of the most consistent trends of the past decade has been the gains of black students relative to nonminority students--a pattern that appears without serious exception on every test identified in this study in which separate data for black students are available. Although this pattern results in part from the more rapid deterioration of scores among nonminority students during the last years of the decline, much of the relative gain of black students is real, in that it reflects greater subsequent improvement in their performance than has been shown by nonminority students. The gap in average scores between black and nonminority students, however, remains large on most tests. Hispanic students also appear to have gained relative to nonminority students, although the data pertaining to Hispanic students are less clear-cut.

Because various types of schools are influenced by different educational practices and social trends, information about achievement trends in different types of schools also has an important bearing on explanations of the trends. It is therefore striking that test scores declined among students in Catholic schools in the United States and Canadian schools as well. In contrast, the existing data, though very sparse, suggest that trends in two other categories of schools--those with high concentrations of minority students and those located in disadvantaged urban communities--have diverged markedly from national trends in recent years. Schools in both categories appear to be gaining appreciably relative to the national average.

The Average Level of Performance on Tests

Despite the recent rise in test scores, the average performance among certain groups and, in some instances, nationwide remains distressingly low. Recent National Assessments of Educational Progress (NAEP) in reading, writing, mathematics, and literacy are rife with illustrations of important skills that large segments of the student population are failing to master. These deficiencies are particularly clear in the assessments of high school students and young adults.

The National Assessments of mathematics, for example, indicate that many students are failing to master even fairly rudimentary skills, particularly when they must reason for themselves what skills to apply rather than simply use a specified arithmetic algorithm. Among 17-year-olds still enrolled in school, only 50 percent to 60 percent (depending on the year of the assessment) were able to solve simple problems involving percentages. (An example is the question: "A hockey team won 5 of its 20 games. What percent of the games did it win?") The proportion able to calculate the cost of electricity per kilowatt hour, given a highly simplified electrical bill, varied from 5 percent to 12 percent, again depending on the year. ^{4/}

The National Assessment of literacy conducted in 1986 revealed striking deficiencies in the ability of young adults (ages 21-25) to use written text in a variety of ways. ^{5/} Less than 40 percent, for example, could synthesize the main argument of a lengthy newspaper article. ^{6/} Roughly 60 percent could extract information from a bar graph, use a chart to pick an appropriate grade of sandpaper, or follow directions using a street map. Given the disturbing level of performance in the mathematics assessments, it is not surprising that some items in the literacy assessment that entailed the use of arithmetic also revealed serious deficiencies. One question presented a simple menu and asked respondents to answer two questions: how much change they would get from a given amount of money

-
4. National Assessment of Educational Progress, *Changes in Mathematical Achievement, 1973-1978* (Denver: NAEP/Education Commission of the States, 1979). These data, which reflect tests administered in both 1973 and 1978, are among the most recent nationally representative data about the mathematics achievement of 17-year-old students. Although current mathematics achievement is probably appreciably higher than that of 1978, it is not likely to be dramatically higher than that of 1973, which was roughly six or seven years before the end of the decline in that age group.
 5. The NAEP literacy assessment differed from that of reading in three important respects: the literacy assessment considered a far broader range of skills (including, for example, the ability to apply rudimentary arithmetic operations in solving problems presented in written text); it tested older youths (ages 21-25, rather than ages 9, 13, and 17); and it included in the sample youths who had dropped out of school.
 6. The proportions of tested individuals noted here as showing a given skill are only approximate. In contrast to many of the earlier National Assessments, the literacy results were not reported in terms of the proportion responding correctly to specific test items. Rather, the proportion performing at a given level of proficiency was reported, along with one or two items indicative of the skills required to demonstrate that level of proficiency. The proportion responding correctly to one of the illustrative test items would generally be slightly different from the proportion showing that level of proficiency, based on all relevant items.

if they ordered two specific items, and how much would be required for a 10 percent tip. Only about 40 percent correctly answered both questions.

IMPORTANT GAPS AND INCONSISTENCIES IN THE ACHIEVEMENT DATA

Examination of a broad array of achievement tests adds considerably to the information that can be obtained from any single test, even if that test yields data of particularly high quality. Yet examining the available test data also reveals the limits of what is currently known about educational achievement. A number of important questions are simply not adequately addressed by available data, and some conclusions that appear straightforward in a single source of achievement data are shown to be questionable when many sources are considered. These gaps and inconsistencies in the data are important not only for understanding the condition of education, but also for explaining recent trends; some of the common explanations are based on aspects of the recent trends that are striking in the results of one or two tests but fail to appear--or are contradicted--in the results of others.

The inconsistencies in the existing test data affect even some of the most fundamental conclusions about recent trends. For example, the size of the decline differed substantially among tests. Tests have also offered dramatically dissimilar pictures of relative trends among different subject areas--an important pattern for explaining the trends, because many explanations are based on factors that would affect some subjects more than others. Regional differences in trends have also varied among tests: the National Assessments have tended to show more favorable trends in the South, which is by some measures the lowest-scoring region, than elsewhere. On the other hand, the only other nationally representative study of regional disparities in trends indicated that declines in scores among high school seniors in the South ranged from being comparable to those elsewhere in one subject to being far worse in another. ^{7/}

-
7. These latter results reflect a comparison of the National Longitudinal Study of the High School Seniors Class of 1972 and the High School and Beyond study. See Rock and others, *Factors Associated With Decline of Test Scores*, Appendix D. Rock used standard Census definitions of the regions, while the National Assessment included in other regions several states that the Census classifies as part of the South. The difference between the results of the two studies is so large, however, that it is very unlikely that this discrepancy in definitions could account for it.

It is also uncertain from the available data whether the trends in test scores vary consistently among achievement subgroups--that is, among groups differing in their initial levels of achievement. Relatively favorable trends among low-achieving students appeared clearly in the NAEP and have figured prominently in some explanations of recent trends. When one considers a variety of tests, however, the information on relative trends among achievement subgroups appears to be a welter of inconsistent findings and disparate definitions of groups. Moreover, comparison of trends among achievement subgroups is hindered by a number of serious technical obstacles. The use of alternative (and equally defensible) methods of scaling and reporting test scores, for example, can fundamentally alter the conclusions one reaches, and the published data are insufficient to sort through the resulting confusion.

Also unanswered is the question of whether the recent rise in scores is beginning to falter. The data offer little reason to doubt that scores in the higher grades will continue rising for several years as the cohorts that recently produced gains in the lower grades progress through school, just as earlier gains in the lower grades were echoed later in the higher grades. Any number of factors could deflect those trends--either augmenting the gains or lessening them--but the data as yet do not indicate such a change (see box on facing page). In contrast, some achievement tests have shown stable scores in the early grades during the past few years, while other tests have shown continuing gains. Only the accumulation of additional information over the next few years will clarify whether progress in the lower grades has indeed ceased for the time being and, if so, whether that stagnation will be duplicated in the higher grades as the affected cohorts progress through school.